Research on Humanities and Social Sciences
ISSN (Paper)2224-5766 ISSN (Online)2225-0484 (Online)
Vol.5, No.5, 2015

Using Muraki Model in Selecting Items for an E-assessment Questionnaire of Faculty Members at Al al-Bayt University

Dr. Iyad Mohammed Hamadneh
Faculty of Educational Sciences – Al al-Bayt University
Director of Faculty Development Center
Mafraq- Jordan

Abstract

This study aimed at using Muraki generalized partial credit model in selecting items for an electronic assessment questionnaire, which asses faculty members performance at Al al-Bayt University, by using the responses of (348) students during the second semester of the academic year 2013/2014 in e-assessment consisting of (20) items. Students responses were analyzed according to the expectations of Muraki generalized partial credit model using statistical software (Winsteps, version 2.88; SPSS). Results of the analysis showed items' fitting for all the items of an E-assessment Questionnaire with the expectations of the model, the reliability of the questionnaire has reached (0.97). In light of the findings of the study, the researcher recommends using a questionnaire consisting of (20) items to assess the performance of faculty members at Al al-Bayt University, and the need to be translated into English, and included instructions for the appropriate response to it.

Keywords: E-assessment, Item Response Theory, Muraki Generalized Partial Credit Model, Faculty Members Performance.

Introduction

Faculty member at the University is major axis of success the learning process; because of the individual qualification adequacy and tender to carry out his duties towards knowledge, teaching, research and Service University and the community in a high quality of performance style. This is traditionally a letter to assess the performance of faculty members to contribute the realization of the principle of excellence in the performance of the university mission, through: Follow-up performance in areas that include the duties and functions, revealing the strengths and weaknesses in his performance in preparation for planning specialized training courses, and linking scientific upgrade academic performance destruction.

In line with this trend, Faculty Development Center (FDC) at Al al-Bayt University was established in 2003/2004 to achieve the goals of the University through developing the faculty members' skills in domains of teaching, evaluation, research, tutorship, and guidance in line with current scientific and technological challenges.

E-assessment is the process of using information networks, and computer equipment and educational software using assessment methods to collect and analyze student responses (Maria, 2010). E-assessment according to Mina (2010) is the process to get the exact description of the data and provide useful information to judge the decision alternatives, and provide safety measures for each to preserve the confidentiality and privacy.

The identification of the electronic assessment are replicas of traditional paper-based questionnaires, but they lead by a computer, where the student read the items from the screen directly and answer them via traditional input tools such as a mouse or keyboard or screen (Jordan & Mitchell, 2009; Simos, 2009).

Shepard study (2009) was aimed to explore an e-assessment of the difficulties of measuring learning outcomes and to overcome the assessment difficulties using traditional methods. The study suggested the need for diversity in the ways of practical e-assessment to provide an opportunity for students to develop thinking and learning to achieve the highest level and quality in higher education skills.

McCann (2010) discussed Factors affecting the adoption of e-assessment system, and the results showed that E-assessment contributed to improve student learning and curriculum revision and change the education of students of culture, and the need to receive faculty members' adequate training on the electronic assessment system.

Since the emergence of interest in university teaching assessment, the studies suggest that students and teachers are the appropriate source for the selection of items containing the practices of teaching; because they are the parties involved in such practices (Lesser & Ferrand, 2000). Fadi & Barbara (2002) pointed that the participation of students in assessing the teaching performance is still a controversial issue, even at universities where assessment and student participation process have reached an advanced stage, not because of the principle of participation, but also how to use the results of the assessment. Where Marlin (1987) saw that students

perceive assessment opportunity to unload their stress, and venting their feelings, also Jacobs (1987) confirmed that 40% of students conspire to impact negatively on the teacher, and assembling greater what can be of low estimates. Whereas Seldin (2005) saw that the credibility of e-assessment can be doubted if the percentage of responders' students less than 75% of those registered in the course.

The estimates of students can be affected by many factors and variables that can contribute to cumulative errors. There are many studies on factors that may affect the sincerity of students estimates of a faculty member, including the nature of the course being mandatory or optional (Lesser & Ferrand, 2000), and teaching experience (Rayder, 1988), and academic rank (Gray & Brandenburg, 1985), sex and nationality (Stack, 2003), and the date of the assessment process (Sixbury & Cashin, 1995), and the type of college (humanitarian, scientific) (Tweissi & Samarah, 2014), and inflation marks (Germain & Scandura, 2005).

Researchers in the field of psychological and educational measurement do great efforts in order to build and develop tests that can measure the capacity of individuals accurately and objectively, as a result of these efforts, psychological and mental measurement field has seen significant developments related to the construction and analysis of the psychological and mental achievement tests (Loyd, 1988).

Among these efforts is the applications of (Item Response Theory, IRT), which is also called (Latent Trait Models) due to the presence of different models within the framework of this theory, vary depending on the assumptions regarding the experimental data, it is assumed that the items vary in difficulty or in their ability to distinguish between different levels of ability, or that the correct answer in their is affected by random guess (Gruijter & kamp, 2005).

Muraki generalized partial credit model is one of IRT models, which you can estimate item parameters, difficulty parameter and a discrimination parameter, with Polytomous responses. For Muraki generalized partial credit model to be objective measurement, it must be separated items parameters levels (Item-Free) about the capabilities of individuals who answer them (Person-Free), as well as the separation of Person-Free for Item-Free, these parameters can be separated using Unconditional Maximum Likelihood Procedure, a so-called "Specific Objectivity" (Muraki, 1992). Detection Item-Free and Person-Free for the expectations of this model, and get to know psychometric properties of the E-assessment Questionnaire can be done Using special computer software for Muraki model, such as statistical software (PARSCAL) (Muraki & Bock, 1997), or (Winsteps, version 2.88) (Lincare, 2002).

Problem and Questions of the Study

The great significance of (FDC) lies in supporting , enhancing, and achieving the principles of distinguish for a faculty member, who is considered the core of teaching-learning process, due to his/her experience, and knowledge transferring. A faculty member is considered responsible for students' thoughts formation, for student's behaviors, for thinking up their traditions and apotheosis, and for merging them with the society they live in as well. Thus, any change or improvement must start with a faculty member.

The use of technology in educational institutions is a requirement of accreditation requirements, and decision-makers are in race to invest technology in the educational process. There is no doubt that the evaluation as an integral part of the educational process is not in isolation from the trend toward technology (Blatchford et al., 2006; Khezi, 2010).

On the premise that the objective measurement tool is the spirit of the correct assessment process, this study came in an attempt to develop an E-assessment Questionnaire (scale) for the performance of faculty members, so agree with measurement objective criteria according to (IRT), which sees Anstasi & Urbina (1997) as constituting the current and future framework for the development of scales.

Based on my knowledge, the development of tools to measure the performance of faculty members did not receive interest of Jordanian researchers, Confined mostly to translate some measures. This study is attempted to develop an electronic scale measures the performance of faculty members at Al al-Bayt University by answering the main question: "What is the possibility of developing an E-assessment Questionnaire of the performance of faculty members accordance with Muraki generalized partial credit model "?

Specifically, the current study seeks to answer the following two questions:

1- What is the degree of E-assessment questionnaire data fitting with Muraki model?

2- What semantics of reliability and validity are available for each item of the E-assessment questionnaire?

3- What is the relative efficiency of information in the E-assessment questionnaire?

Aims of the study

This study aims to develop a scale (E-assessment questionnaire) to assess the performance of faculty members, and assess the psychometric properties of the scale, according to Muraki generalized partial credit model, And verification of appropriate the scale for using and practicality application, achieving an objective measurement requirements; so it can be used by developing the performance of faculty members Center (FDC) at Al al-Bayt University.

Importance of the study

The importance of the current study comes in response to the growing interest in assessing the performance of faculty members, by providing a scale (E-assessment questionnaire) which has a good degree of objectivity and accuracy. It is also important because of using Muraki generalized partial credit model as one of (IRT) models in the development of psychological and educational scales, and that the scarcity of Jordanian studies used this model in the development of scales; to conduct subsequent studies which highlight the various applications of this model in psychological and educational measurement.

Limitations of the Study

- The study sample was limited to students in the first year of the first undergraduate (BA) at the University of Al al-Bayt during the second semester of the academic year 2013/2014.
- The study was limited by e-assessment questionnaire that use in (FDC).
- The study was limited by using Muraki model according to (IRT).

.

Methodology of the Study

Participants

The population of the study consisted of all undergraduate students at Al al-Bayt University during the second semester of the academic year 2013/2014, where it was applied e-assessment questionnaire on 16396 students of the total number of university students 17833 students, divided into 14 College, Institute and the Centre; to solicit their views on the performance of the faculty members level 378 members, and by general assessment of %92, has an average overall assessment of faculty members at the university level body %82.44 (A report on the results of an assessment of faculty members, second semester 2013/2014, Faculty Development Center, Al al-Bayt University, Jordan).

The study sample consisted of (348) students from the level of the first year at the university, was chosen randomly (Cluster sample), and unit of choice was the section.

Instrument

FDC pays a great deal of attention to students' assessment of their faculty member's performance. The center started applying a system of faculty members assessment electronically round the first and second semester academic year, this trial still goes on up to this moment. The performance assessment corresponds in accordance with best teaching-learning practices, it is also reconsidered when there is a need for that, depending on scientific studies outcomes in this domains. Such assessments are kept electronically in a system specified for that. Each staff member has the right to see his/her performance assessment of the semester through his/her personal portal, so can plan for better performance in the future, and being encouraged to perform in accordance with high levels through best concerns and support of University board.

Questionnaire prepared by (FDC) consists of (24) private items and (4) General items, it is provided for students as a hardcopy, and continued to use until the university asked to be developed and made electronically. It has already been done through the study of cash analytical study in private workshops specialists participated in the field of measurement and evaluation, have been found on the experiences of other universities in this area, and look at the literature on the assessment of the performance of teaching in Arab universities faculty members (Yarmouk University; and University of Jordan) and other assessment tools (Mason et al., 2001; Wachtel, 1998; Williams et al., 2000). Based on the foregoing E-questionnaire has been prepared, consisted of (20) items, this E-questionnaire covered a number of themes related to the performance of a faculty member, namely: the area's commitment to the faculty member, and the field of measurement and evaluation, and teaching methods, and the field of teacher dealing with students, according to likert scale quintet (Excellent, Very good, Good, Acceptable, Weak).

Statistical Treatment

- Verifying assuming (Unidimensionality) of the E-questionnaire, using statistical software (SPSS).

- Used statistical software (Winsteps, version 2.88), to analyze the data, and to answer the study questions.

Results and Discussion

First, the results regarding the degree of E-assessment questionnaire data fitting with Muraki model. To answer this question of the study, it was necessary to follow the following steps:

A) – Check the assumption of Unidimensionality

Verification of Muraki model assumptions as one of (IRT) models. After the data of student responses entered into the computer's memory, it has been verified assuming (Unidimensionality), analysis using the factor analysis of (Principal Component Analysis), data identifying the e-assessment-related responses (348) students represent a sample staging (20) items, according to the statistical program SPSS and the values of (Eigenvalue) were calculated, and note the ratio of (Explained Variance) for each factor, and the ratio of Cumulative Explained Variation corresponding to each factor values, as shown in Table (1).

Table (1): Summary of Exploratory Factor Analysis Results for the E-assessment questionnaire (N = 348)

Factor	*Eigenvalues*	% Explained Variance	% Cumulative Explained Variance
The first	11.71	%58.50	%58.50
The second	1.27	%6.30	%64.80
The third	1.32	%6	%70.80

Can be seen from Table (1), that there is one factor only increased its eigenvalue (2), which explains what percentage (%58.5) of the variation subject's signs to identify e-assessment, and reached the eigenvalue to this factor value (11.71), which is very high if compared to the underlying eigenvalues of other factors. This is a strong indication of the Unidimensionality, and noted that the %Explained Variance of second and the third factors are very close, meaning that there are relatively uniform and almost stable in %Explained Variance with the exception of first factor, and this means check Unidimensionality, where a number of researchers, including Recase (Recase as cited in Hattie, 1985) pointed out that, if the %Explained Variance of first factor is the greatest, this index on a Unidimensionality, and Recase selects that value to be at least 20% of the variance.
Depending on the Lord index (Lord, 1980) Special Unidimensionality, which stipulates that the items check Unidimensionality: If the eigenvalue of the first factor relative to the eigenvalue of the second factor is greater than (2).

The assumption of Unidimensionality is reinforced by using what is known as tested (Scree plot) that appears in Figure (1).

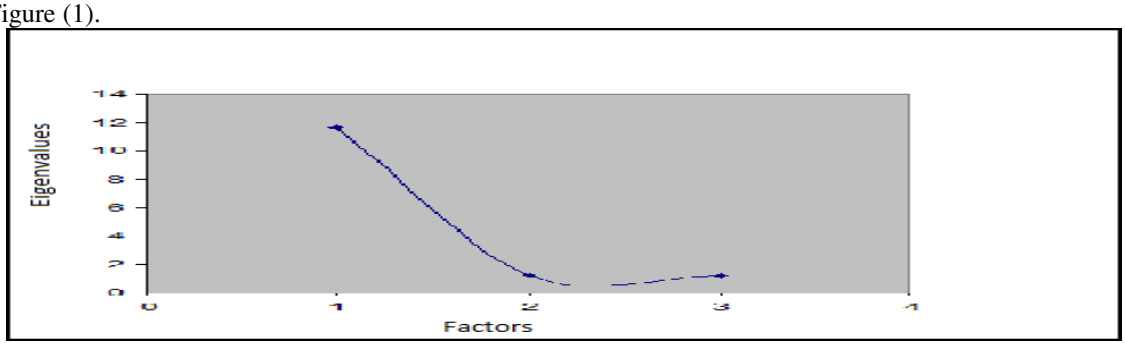

Figure (1): Scree Plot of the eigenvalues for E-assessment questionnaire factors on the total data

Can be seen from Figure (1), that the eigenvalue of the first factor is characterized by clearly on the rest factors eigenvalues, and this is also an indication of a Unidimensionality data.

B) - Check the assumption of (Goodness - of - Fit) for the data of the current study

To determine the degree of E-assessment questionnaire data fitting with Muraki model, the execution of orders using the program file (WINSTEPS, VERSION 2.88) designed for this purpose. Using the method of (Marginal Maximum Likelihood, MML-EM).

To see the (Person Fit), ability of each person in addition to the standard error in measurement of this ability has been estimated, and statistically (INFIT & OUTFIT), and Mean Square Statistic (MNSQ), and the Standardized Information Weighted Fit Statistics for Persons (ZSTD) have been estimated, As shown in Table (2).

Table (2): Mean and Standard Deviation for each person Ability and Standard Error, and INFIT and OUTFIT (N= 348, Items= 20)

Descriptive Statistic	Ability	Standard Error	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	0.05	0.29	1.04	-0.1	1.03	-0.2
Standard Deviation	1.12	0.07	0.49	1.3	0.69	1.2

Table (2) shows the approach of the mean of MNSQ from one, which is the ideal situation as expected by the model, also notes that the statistical value ZSTD of INFIT has reached (-0.1), the standard deviation has been reached (1.3) which is close to the ideal values assumed by the model, namely (0, 1) respectively, also notes that the statistical value ZSTD of OUTFIT has reached (-0.2) and standard deviation equal to (1.2) and is also close to the ideal values assumed by the model, namely, (0, 1) respectively.

Then, when examining the statistical values of OUTFIT for persons it shows the presence of (26) persons their responses away from the expected responses depending on their abilities, in the sense that MNSQ values corresponding of their abilities larger than one, theses are the values which are expected by the model, or the statistical values of MNSQ corresponding to the abilities larger than (+2). As indicated by the program these persons represent more responses MISFIT to the model (Most Miss fitting Response Strings), as shown in Table (3).

Table (3): Persons Numbers (arranged) whom MISFIT the Model, with their MEAN-SQUARE & OUTFIT (N= 26)

PERSONNUMBER	MEAN-SQUARE	OUTFIT	PERSONNUMBER	MEAN-SQUARE	OUTFIT
284	9.03	4.9	290	2.05	2.2
11	5.68	1.2	268	2.03	1.5
220	3.73	2.9	331	1.84	2
262	3.71	3.1	182	1.37	0.6
3	3.32	4.7	144	2.1	2.5
333	2.21	1.6	326	2.08	2.7
168	3	2.8	292	1.88	2.3
296	2.84	3.2	334	1.52	1.5
282	1.64	1	255	1.81	2.1
321	1.64	1	135	1.79	2.1
19	2.62	3.8	88	1.79	2
310	1.13	0.2	270	1.78	2
321	1.13	0.2	347	1.22	0.4

Also, Table (4) contains a sketch expressed. As well as PERSON FIT GRAPH and the location of each member of this order, so as to INFIT MEAN-SQUARE and OUTFIT MEAN-SQUARE.

Table (4): PERSON FIT GRAPH: MISFIT ORDER
(N= 26)

PERSON FIT GRAPH: MISFIT ORDER

	ENTRY	MEASURE	INFIT MEAN-SQUARE	OUTFIT MEAN-SQUARE	
NUMBER	- +	0 0.7 1 1.3 2	0 0.7 1 1.3 2	PERSON	
284	*	: . : *	A : . : *	10284	
11	*	: . :*	B : . : *	10011	
220	*	: . : *	C : . : *	10220	
262	*	: . : *	D : . : *	10262	
3	*	: . : *	E : . : *	10003	
333	*	: . : *	F : . : *	10333	
168	*	: . : *	G : . : *	10168	
296	*	: . : *	H : . : *	10296	
282	*	: . : *	I : . : *	10282	
321	*	: . : *	J : . : *	10321	
19	*	: . : *	K : . : *	10019	
310	*	: . : *	L : .* :	10310	
321	*	: . : *	M : .* :	10320	
290	*	: . : *	N : . : *	10290	
268	*	: . : *	O : . : *	10268	
331	*	: . : *	P : . : *	10331	
182	*	: . : *	Q : . *	10182	
144	*	: . : *	R : . : *	10144	
326	*	: . : *	S : . : *	10326	
292	*	: . : *	T : . : *	10292	
334	*	: . : *	U : . : *	10334	
255	*	: . : *	V : . : *	10255	
135	*	: . : *	W : . : *	10135	
88	*	: . : *	X : . : *	10088	
270	*	: . : *	Y : . : *	10270	
347	*	: . : *	Z : . *:	10347	
28	*	: . : *	: . : *	10028	
385	*	: . : *	: . : *	10385	
-OMIT-					

Can be seen from Table (4), that the majority of persons MISFIT have INFIT MEAN-SQUARE and OUTFIT MEAN-SQUARE corresponding to their abilities were outside the fit order (0.7- 1.3).

After excluding the twenty-six persons (Misfit the Model). We re-analyzed to test (Item Fit) of the expectations of Muraki model, as shown in Table (5).

Table (5): Items numbers and values of their OUTFIT and INFIT and measures and Point Biserial correlations (PTBIS)

ITEMS STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS	
12	481	322	0.33	0.07	1.25	3.3	1.22	2.3	A0.34	2.7	I12
16	577	322	0.59	0.07	1.18	2.3	1.19	2.3	B0.41	2.7	I16
17	411	322	1.19	0.06	1.17	2.1	1.11	1.0	C0.41	3.1	I17
18	314	322	1.59	0.07	1.17	1.9	1.16	1.3	D0.36	1.1	I18
20	371	322	1.50	0.06	1.14	1.9	1.15	1.2	E0.42	3.7	I20
14	646	322	0.21	0.07	1.07	1.0	1.05	0.6	F0.48	1.1	I14
10	596	322	0.47	0.07	1.06	0.8	1.06	0.8	G0.47	4.1	I10
19	513	322	0.75	0.06	1.06	0.8	1.03	0.3	H0.50	3.1	I19
15	615	322	0.56	0.08	1.05	0.7	1.04	0.5	I0.47	4.1	I15
9	665	322	0.07	0.08	0.97	-0.4	0.97	-0.4	J0.52	3.1	I9
5	750	322	-0.22	0.08	0.97	-0.4	0.95	-0.6	j0.55	1.1	I5
2	842	322	-0.82	0.07	0.91	-1.2	0.96	-0.5	i0.59	1.1	I2
3	404	322	1.36	0.06	0.94	-0.8	0.90	-0.9	h0.54	1.1	I3
11	1000	322	-1.18	0.08	0.94	-0.7	0.92	-0.9	g0.59	1.1	I11
1	1043	322	-1.00	0.08	0.82	-2.6	0.90	-0.9	f0.64	.7	I1
4	1043	322	-1.49	0.08	0.89	-1.3	0.83	-1.7	e0.63	2.7	I4
6	977	322	-1.17	0.07	0.86	-1.9	0.82	-2.0	d0.63	1.7	I6
13	947	322	-0.94	0.08	0.86	-1.7	0.84	-1.7	c0.62	2.1	I13
7	939	322	-0.99	0.07	0.82	-2.5	0.81	-2.3	b0.66	3.1	I7
8	875	322	-0.82	0.07	0.81	-2.8	0.79	-2.8	a0.66	1.7	I8
MEAN	700.	322.	0.00	0.07	1.00	-0.1	0.99	-0.2			
S.D.	237.	0.	0.97	0.01	0.13	1.8	0.13	1.4			

Table (5) shows that the statistical values of items INFIT were all non-statistically significant, and this indicates that the agreement between the curve observed and the best curve to matches the model, which means that there is independent of items difficulty about the sample, and then stable for this difficulty across different levels of ability. As well as showing there were no statistically significant differences in the statistical values of items OUTFIT and this means that the items all measure one trait, and the values of point biserial correlation coefficient (discriminations), were all positive and to some extent high and tight, and this shows that the all the items in E-assessment questionnaire- used in the current study- have the ability to distinguish between persons.

Also, Table (6) shows a sketch expressed for an Items FIT GRAPH and the location of each Item of this order, so as to INFIT MEAN-SQUARE and OUTFIT MEAN-SQUARE.

Table (6): Items FIT GRAPH: MISFIT ORDER
(Items= 20)

ITEMS FIT GRAPH: MISFIT ORDER

ENTRY NUMBER	MEASURE −	+	INFIT MEAN-SQUARE 0 0.7 1 1.3 2	OUTFIT MEAN-SQUARE 0 0.7 1 1.3 2	ITEMS	
12	*		: . *:	A : . *:	2.7	I12
16	*		: .* :	B : .* :	2.7	I16
17		*	: .* :	C : .* :	3.1	I17
18		*|	: .* :	D : .* :	1.1	I18
20		*|	: .* :	E : .* :	3.7	I20
14	*		: * :	F : * :	1.1	I14
10	*		: * :	G : * :	4.1	I10
19	*		: * :	H : * :	3.1	I19
15	*		: * :	I : * :	4.1	I15
9	*		: *. :	J : *. :	3.1	I9
5	*		: *. :	j : *. :	1.1	I5
2	*		: *. :	i : *. :	1.1	I2
3		*	: *. :	h :* . :	1.1	I3
11	*		: *. :	g : *. :	1.1	I11
1	*		:* . :	f : *. :	.7	I1
4	*		:* . :	e :* . :	2.7	I4
6	*		:* . :	d :* . :	1.7	I6
13	*		:* . :	c :* . :	2.1	I13
7	*		:* . :	b :* . :	3.1	I7
8	*		:* . :	a * . :	1.7	I8

In light of these indicators, it is shown that all items had MEAN-SQUARE values confined within the fit limits (0.7) minimum and (1.3) maximum, and since the values of these statistics were within the range, it means that there are no significant statistical differences, and therefore no item of the electronic- assessment questionnaire was deleted.

Finally, Figure (2) Summarizes, a map of scaling persons and items after verification the data FIT to the expectations of Muraki model. Figure (2) shows that the E-assessment questionnaire used to measure the extent of acceptable capacity (-1.48- 1.58) and measuring one trait (unidimentionality) feature, which reflects the validity of the construction of the questionnaire.

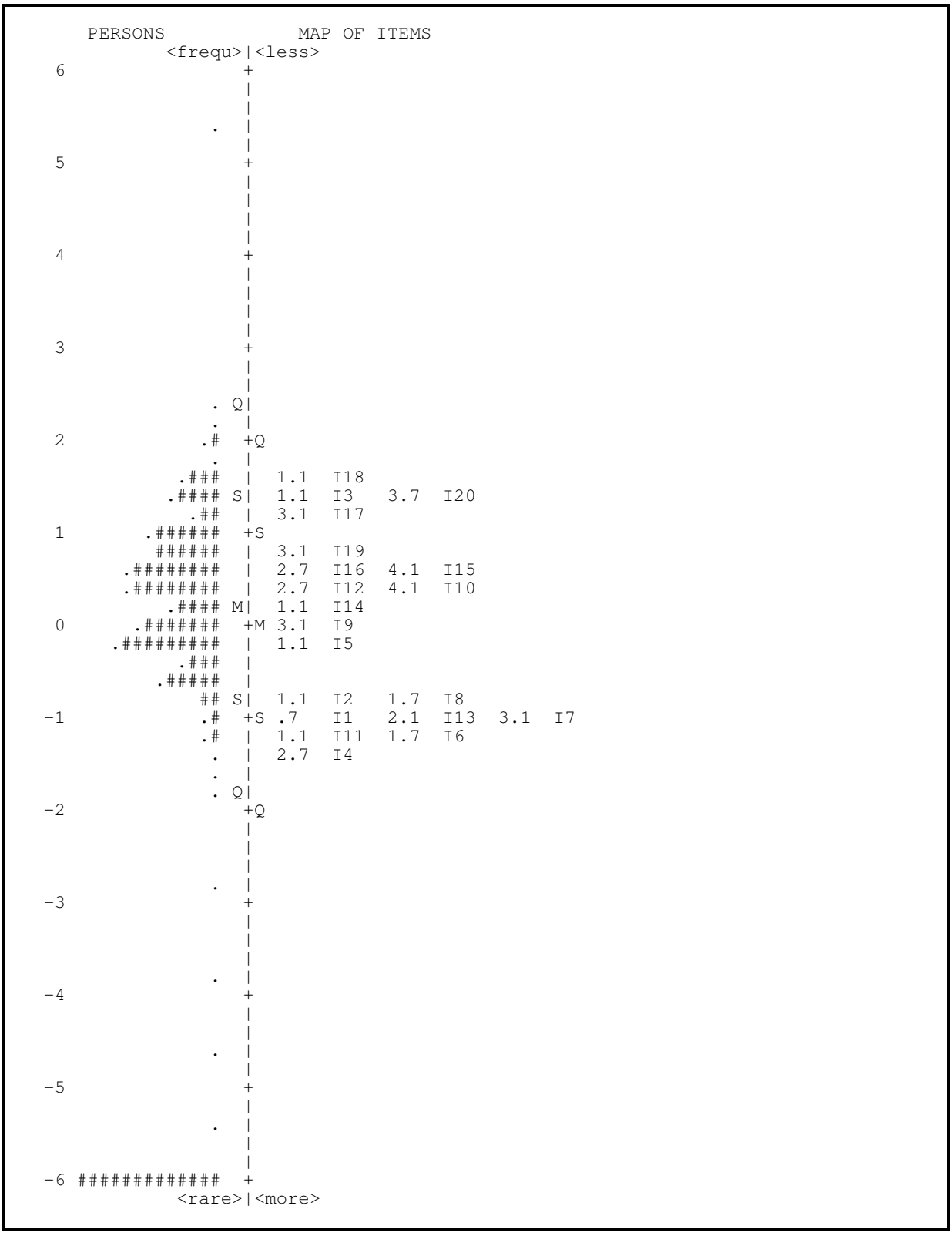

FIGURE (2): MAP OF SCALING PERSONS AND ITEMS

Secondly: The results related of reliability and validity are available for each item of the E-assessment questionnaire free from the characteristics of persons.

Hambleton & Swaminathan (1985) Explain that the use of IRT models of measurement in developing psychological and educational instruments, requires the provision of evidences of instruments validity, because the data FIT (PERSONS FIT) and (ITEMS FIT) to the expectations of the model used in the development of the instrument does not mean the availability of instrument validity.

Coincided with the development of E-assessment questionnaire meeting called by the Tafila Technical University to directors of faculty development centers in Jordanian universities dated 25 December, 2014, a regular meeting held quarterly at several universities in order to discuss ways to improve the performance of faculty members in Jordanian universities. Took advantage of the researcher the opportunity to check out the content validity of the scale (E-assessment questionnaire used at Al al-Bayt University) viewing on arbitrators specialists with expertise and efficiency numbered (11) arbitrator, all directors occupy the post of Director of FDC in their universities, six of whom specialize in Measurement and Evaluation and the rest of the campaign doctorate in philosophy and curriculum and teaching methods; and to take their views on the paragraphs, and propose what they see as an appropriate amendment.

Where the study questionnaire cash analytical study, by reading every paragraph (Item), and trial in extensive discussion takes into account the meaning of the phrase and how it is understood by the students.

In light of the arbitrators' estimates and the researcher discussion with them, some paragraphs of the amendment to the scale. The arbitrators unanimously that there are no paragraphs very difficult nor too easy paragraphs, and paragraphs were suitable for those who are examined. The language of two paragraphs has been modified on which the whole arbitrators that drafted the form in which they suggested probably best. all the judges stressed the need to include a measure on the occasion of the student instructions to respond to it in earnest, and suggested the need to translate the paragraphs of the scale to the English language like the scale of ALYarmouk University, and the University of Jordan, under the pretext of the existence of some students of other nationalities are required to respond to all the paragraphs of the scale if limited, and all the judges stressed to be directed by the scale in Arabic and English on the same scale, without being deleted a clause.

With respect to the scale (E-assessment questionnaire) reliability, Item Separation Index was around (13.10) and Person Separation Index reached (3.22), in the sense that the scale reliability has reached (0.96) and the reliability of persons has reached (0.90), a undoubtedly very high values indicate the adequacy of the paragraphs (items) in the assessment questionnaire separation between persons and the distinguish between different levels of persons ability, and definition items related attribute that measured by these items (Wright & Masters, 1982).

There is a statistical indicator else can be referred to which is called (Ability Strata) where it has reached (4.63), while the value of Items Strata has reached (17.82). According to these values the e-assessment questionnaire items have an ability to show the individual differences between persons in terms of the degree of trait ownership.

Third: The results concerning the (Relative Efficiency of Information) in E-assessment questionnaire.

To find the relative efficiency of the e-assessment for participants of the study sample, Information Function were found, all items have been introduced in the analysis process, especially that of (20) items, as well as the selection of 200 persons respondents their ability divided to connected to the ability of lower-ability (-5.32) Logit to higher- ability (5.44) Logit. Then the data were entered in the computer memory and using the program (WINSTEPS, VERSION 2.88) to estimate the Information Function (IF) at selected levels of ability estimated. As shown in Table (7).

Table (7): Ability estimates and Information's Function (IF) values that provided by the E-assessment questionnaire at selected levels of ability

Ability	IF	Ability	IF	Ability	IF	Ability	IF
-8.98	0.04	-3.55	4.21	1.87	9.73	7.33	0.2
-8.66	0.06	-3.22	5.18	2.22	8.38	7.66	0.15
-8.32	0.08	-2.88	6.3	2.56	7.13	7.99	0.1
-7.98	0.12	-2.54	7.49	2.9	6		
-7.64	0.16	-2.2	8.57	3.24	4.99		
-7.3	0.23	-1.86	9.39	3.58	4.11		
-6.96	0.31	-1.52	9.97	3.92	3.34		
-6.62	0.43	-1.18	10.5	4.26	2.68		
-6.28	0.58	-0.84	11.1	4.6	2.12		
-5.94	0.77	-0.5	11.77	4.94	1.66		
-5.6	1.02	-0.16	12.42	5.28	1.27		
-5.26	1.34	0.18	12.93	5.62	0.96		
-4.92	1.72	0.52	13.14	5.96	0.71		
-4.58	2.19	0.86	12.9	6.3	0.53		
-4.24	2.75	1.2	12.17	6.64	0.39		
-3.9	3.42	1.54	11.04	6.98	0.28		

Also, Figure (3) shows a graph of the relationship between the information function values of E-assessment questionnaire at every level of ability levels values.

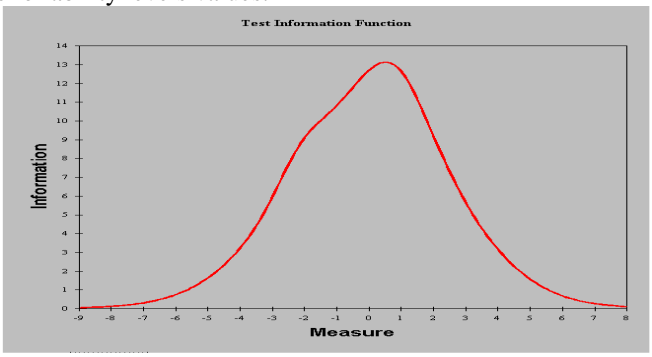

Figure (3): E-assessment questionnaire Information Function

Can be seen from Table (7), and Figure (3) that E-assessment questionnaire Information Function was the largest as possible at zero Logit ability level. In the sense that the E-assessment questionnaire gives a more effective information when persons with medium ability, while the information function provided by the E-assessment questionnaire of values were minimal at high and low levels of ability, in the sense that the E-assessment questionnaire gives little information when persons with high and low abilities.

It should be noted that the scale Information Function(IF) one of the indicators that inferred the reliability of the scale in (IRT), where the curve of IF works unlike the Standard Error, and therefore increase the amount of IF leads to decrease the standard error of measurement. So it is expected that the greater of scale (IF) at a certain level of ability increased reliability, in the sense that at least the standard error of measurement, which provides the opportunity to estimate standard error at every level of ability levels, and see how the contribution of each item in determining the measurement accuracy, So these items have acceptable degree of validity and reliability (Reeve & Fayers, 2004) .

Recommendations

In light of the current study findings, the researcher recommends the following:

1- Taking the finding of the current study, in terms of the need to provide Al al-Bayt University E-assessment questionnaire with accurate instructions, and the need to translate all its items- (20) items that have acceptable degree of validity and reliability- into English language as well as Arabic language; that some respondents it may be of foreign students and non-Jordanian nationalities, who are not fluent in understanding Arabic language or content of the items.

2- Take advantage of Al al-Bayt Universityye E-assessment questionnaire related to a group of 20 items, which cover a wide range on the online feature in the performance of Faculty Members assessment, and

try to make a study according to different statistical software, as well as try to Fit the data to the expectations of the other (IRT) models, and a comparison between them.

References

Anastasi, A., & Urbina, S. (1997). Psychological Testing (7th Ed.). New Jersey: Prentice- Hall, Inc.

Blatchford, P., Baines, C., Rubie, D., Bassett., & Chowne, A. (2006). The effect of a new approach to group work on pupil- pupil teacher- pupil interactions. Journal of Education Psychology, 98(4), 749- 764.

Fadi, N., & Barbara, F. (2002). Faculty views of student evaluation of college teaching. Assessment and Evaluation in Higher Education, 27 (2), 187- 197.

Germain, M., & Scandura, T. (2005). Grade inflation and student individual differences as systematic bias in faculty evaluation. Journal of Instructional Psychology, 32(1), 58- 67.

Gray, D., & Brendenburg, D. (1985). Following student ratings over time with a catalog-based system. Research in Higher Education, 22(2), 156- 168.

Gruijter, D., & Kamp, L. (2005). Statistical Test Theory for Education and Psychology. Retrieved December 30, 2005 from: www.leidenuniv.nl /~ gruijterdnmde.

Hambleton, R., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer Nijhoff Publishing.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research, 19(1), 49-78.

Jacobs, L. (1987). University faculty and students' options of student ratings. Indiana Studies in Higher Education, Report No.55 Bureau of Evaluation and Testing, Indiana University.

Jordan, S., & Metchell, T. (2009). E- assessment for learning? The potential of short- answer free text question with tailored feedback. British Journal of Education Technology, 40 (2), 371- 385.

Lesser, D., & Ferrand, J. (2000). Effect at class size, grades given and academic field on student opinion of instruction. Community College Journal of Research and Practice, 24(4), 269- 278.

Lincare, J. (2002). A user's guide of Winsteps Minsteps Rasch-Model. Computer Program, Chicago. 2002.

Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale, N .J: Lawrence Erlbaum.

Loyd, B. (1988). Application of item response theory for the measurement practitioner. Applied Measurement in Education, 1(2), 135- 143.

Maria, F. (2010). E- assessment within the bologna paradigm: evidence from Portugal. Assessment & Evaluation in Higher Education, 35(7), 819- 830.

Marlin, J. (1987). Student perception of end- of course evaluations. Journal of Higher Education, 58, 704- 716.

Mason, K., Edward, R., & Roach, D. (2001). Student evaluation of instructors: A measure of teaching effectiveness on something else?. Journal of Business Administration, online, 1 (2), 8p, (from: http:// jbao.atu.edu/ mason_ yoach_ htm, retrieved: 20/10/2014).

McCann, A. (2010). Factors affecting the adoption of e- assessment system. Assessment & Evaluation in Higher Education, 35(7), 799- 818.

Mina, C. (2010). Embedded formative e- assessment: who benefits, who falters. Educational Media International, 47(2), 153- 171.

Muraki, E. (1990). Fitting a polytomous item response model to likert type data. Applied Psychological Measurement, 14, 60- 71.

Muraki, H. (1992). A generalized partial credit model: application of an EM algorithm. Applied Psychological Measurement, 16, 159- 176.

Muraki, H., & Rock, R. (1997). PARSCAL version 4.1: IRT item analysis and test scoring for rating scale data (Computer Program). Chicago, IL: Scientific Software.

Rayder, N. (1988). College student ratings of instructors. Journal of Experimental Education, 37, 107- 116.

Reeve, B., & Fayers, P. (2004). Application of Item Response Theory (IRT) Modeling for Building Evaluating Questionnaires Measuring Patient-Reported Outcomes [Online]. Available: http://www.researchgate.net/puplication.

Seldin, P. (2005). The use and abuse of student ratings of professors [Online]. Available: http://www.tre.ucdavis.edu.

Shepard, K. (2009). E is for exploration: Assessing hard to measure learning outcomes. British Journal of Education Technology, 40(2), 386- 398.

Simos, B. (2009). Accessibility in e- assessment. Assessment & Evaluation in Higher Education, 34(3), 293- 303.

Sixbury, G., & Cashin, W. (1995). Description of database for the IDEA diagnostic. From IDEA paper no. 9.

Stack, S. (2003). Research productivity and student evaluation of teaching in social classes. A research note.

Tweissi, A., & Samarah, N. (2014). Attitudes and Satisfaction of Teaching Faculty Members at Mu'tah-University towards the Assessment of Their Teaching Performance by their Students. An-Najah University Journal for Research – Humanities, 28(1), 1- 30.

Wachtel, H. (1998). Student evaluation of college teaching effectiveness: a brief review. Assessment and Evaluation in Higher Education, 23 (2), 191- 212.

Williams, J., Diann, M., & Howard, E. (2000). How faculty evaluation are used in Texas community college. Community College Journal of Research and Practice, 24, 169- 179.

Wright, B., & Master, G. (1982). Rating Scale Analysis. Rasch Measurement (1st Ed.). Chicago: MESA Press.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar